DOCUMENT RESUME

ED 051 861 LI 002 921

TITLE	National Science Research Data Processing and Information Retrieval System; Hearings before the General Subcommittee on Education ... Ninety-First Congress, First Session on H.R. 8809 ... April 29 and 30, 1969.
INSTITUTION	Congress of the U.S., Washington, D.C. House Committee on Education and Labor.
PUB DATE	69
NOTE	13p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Computer Output Microfilm, *Documentation, *Electronic Data Processing, Federal Legislation, *Information Processing, *Information Retrieval, Information Systems, *Microforms
IDENTIFIERS	New York Times Information Retrieval System, *Scientific and Technical Information

ABSTRACT

Testimony given by Dr. John Rothman, Manager of Information Services, New York Times, before the General Subcommittee on Education ... describes the information retrieval system of the New York Times. He describes how the system will process material efficiently and economically, store it securely and in proper order, and make it available via computer to inquirers. The abstracts will be stored in a computer, with the full text stored in microform in an automated device linked to the computer, and appropriate computer-linked input/output facilities. The statement of Dr. Rothman and Mr. Robert S. November, Director of Library Services and Information Division of the New York Times, gives the background information of the New York Times Information Retrieval System, why it was started and what are the future plans. The purpose of the hearing is to better qualify the subcommittee members to evaluate pending legislation for the establishment of a National Science Research Data Processing and Information Retrieval System. (NH)

NATIONAL SCIENCE RESEARCH DATA PROCESSING AND INFORMATION RETRIEVAL SYSTEM

HEARINGS

BEFORE THE

GENERAL SUBCOMMITTEE ON EDUCATION

OF THE

COMMITTEE ON EDUCATION AND LABOR

HOUSE OF REPRESENTATIVES

NINETY-FIRST CONGRESS

FIRST SESSION

ON

H.R. 8809

A BILL TO AMEND TITLE IX OF THE NATIONAL DEFENSE EDUCATION ACT OF 1958 TO PROVIDE FOR ESTABLISHMENT OF A NATIONAL SCIENCE RESEARCH DATA PROCESSING AND INFORMATION RETRIEVAL SYSTEM

HEARINGS HELD IN WASHINGTON, D.C.
APRIL 29 AND 30, 1969

Printed for the use of the Committee on Education and Labor
CARL D. PERKINS, *Chairman*

U.S. GOVERNMENT PRINTING OFFICE
31-464 WASHINGTON : 1969

SCIENCE AND TECHNICAL INFORMATION RETRIEVAL BILL

WEDNESDAY, APRIL 30, 1969

HOUSE OF REPRESENTATIVES,
GENERAL SUBCOMMITTEE ON EDUCATION OF THE
COMMITTEE ON EDUCATION AND LABOR,
Washington, D.C.

The subcommittee met at 10 a.m., pursuant to recess, in room 2261, Rayburn House Office Building, Hon. Roman C. Pucinski, presiding.

Present: Representatives Pucinski, Bell, Ruth, and Dellenback.

Staff members present: Allan Kiron, technical adviser; John F. Jennings majority counsel; Alexandra Kisla, clerk; and Charles W. Radcliffe, minority counsel for education.

Mr. PUCINSKI. The committee will come to order.

We will proceed. The other members of the committee are on their way here and they will be assembling. One of the problems we have this morning is that the majority leader is testifying before the Subcommittee on Poverty and many of our members are down there. But in order to expedite this situation here I thought we would proceed at this time rather than wait because I am sure that Dr. Rothman and Mr. November are anxious to get back to New York.

If we may have Dr. Rothman and Mr. November take the stand we will move along and the other members will join us shortly.

Gentlemen, we are very pleased to have you here this morning. Dr. Rothman, as I understand it you are director of information services for the New York Times. You graduated from Queens College and received your masters degree at New York University and your Ph. D. in comparative literature at Columbia University in 1956.

Certainly you bring to this committee a wealth of knowledge on this subject.

Mr. November, we understand you are the director of library services and information division of the New York Times. You received a degree in economics at Harvard in 1958 and you were a Henry Fellow in economics at Kings College in Cambridge, England in 1959.

We are privileged to have you gentlemen join us this morning and tell us something about the capabilities of information retrieval. We have been watching with great excitement the system that you are developing at the New York Times and I think it is something that was long in coming.

The New York Times is recognized around the world as one of the great storehouses of information. I am very pleased to learn that you gentlemen can set up a more orderly system of retrieving that information. I am sure it is going to play a tremendously important role in this country.

I am hopeful that eventually you are going to develop through Intelstat and similar devices, into a world-wide network. I am not quite sure that the American people are aware of the full significance of what you are doing and the impact that it can have on so many aspects of our social and economic and educational and scientific endeavors.

I view your project with great excitement and am anxious to hear your testimony.

I would recommend that you gentlemen proceed in any manner that you wish. You have a prepared statement which will go in the record if you wish at this time and you be the judge of how you want to proceed. As I say, the other members of the committee will join us here in the morning.

(The statement referred to and a news release follow:)

STATEMENT OF DR. JOHN ROTHMAN, MANAGER OF INFORMATION SERVICES, NEW YORK TIMES

For the last three years The New York Times has been working on the development of an information retrieval system which, when put into operation, will make available vast resources of information now hidden in the clipping library and other facilities of The Times with incredible speed, thoroughness and comprehensiveness.

By using modern data-processing techniques and equipment, the system will process material efficiently and economically, store it securely and in proper order, and make it available via computer to inquirers.

The system will comprise abstracts of The Times and other materials stored in a computer, full text of these materials stored in microform in an automated device linked to the computer, and appropriate computer-linked input/output facilities.

Initially the system will serve the news and editorial departments of The Times and eventually a wide range of customers requiring comprehensive, authoritative information.

HOW THE SYSTEM WILL WORK

The heart of The Times Information Bank will be a third-generation, real-time computer (IBM System 360/50) and software combination which will be designed especially for immediate computer-to-user response (time-shared) with a large number of remote terminals. A large direct-access mass storage facility will be included to insure rapid handling of the large data base.

Remote terminals will be tailored for the customer to any one of three models:

Keyboard input with video output.
Keyboard input with high-speed printer output.
Keyboard input with slow-speed printer output.

Linked with the computer will be an automatic device for the storage of microfiche containing images of the actual clippings. This device will be capable of storing and rapidly retrieving the equivalent of 3.5×10^6 pages of newsprint. Other peripheral equipment will include a microform camera at Microfilming Corporation of America, Inc., a wholly-owned subsidiary of The New York Times, for miniaturization of full text; input terminals—probably cathode ray tube and keyboard—and a computer-telephone interface for audio computer-to-user answer service.

It will be possible for the user, through the audio system, to query the computer directly and receive an answer. It will also be possible to use the audio system for placing calls automatically and answering inquiries that have been previously placed.

The computer will actually converse with the inquirer; the computer will talk, the inquirer will use his dialing device. Modular system design will be a paramount consideration to allow for expansion of central facilities as customer demand increases.

HOW THE SYSTEM WILL BE PUT INTO OPERATION

The information retrieval system will be made operational in four overlapping stages or phases:

Phase 1

The computer, software and full text device will be installed at the 43d Street location of The Times. Only a portion of the design memory size will be installed initially with additions as the needs are forecast. Data input, system checkout, and debugging will be completed and followed by limited use of the system by selected individuals to test fully the operational design features of the system.

Phase 2

Remote terminals will be installed in the New York office of The Times and editors, reporters and other personnel will be shown how to operate the retrieval equipment to receive both abstract and full text items stored in the system.

Phase 3

Remote terminals will be installed at other locations of The Times. Personnel at these locations will be taught the proper use of the equipment for rapid recovery of stored data. There will be no capability at these distant terminals for electronic viewing of full text items; full text will be sent to the location from New York or stored at the distant location in microform.

Phase 4

Ultimately, remote terminals will be installed at customer location. Customers will be trained in the operation of the system and will have full inquiry privileges. Output to customers, however, will be limited to abstracts, citations and a subset of the full text items in storage. Non-Times articles, information from early editions, and killed items will not be transmitted to customer terminals.

For customers who anticipate no need for installation of heavy-use terminals, telephone service with direct hook-up to the computer will be available. A no-charge service will supply computer citations of Times references to the inquirer by means of voice recordings.

During peak load periods, the computer may temporarily store an inquiry, then initiate a call to the inquirer, and automatically supply the citations. If the inquirer desires abstracts in addition to the citations, he will signal the system, a computer assistant will enter the conversation, and a fee will be charged for either telephonic recitation or mailing of the information.

WHAT WILL GO INTO THE SYSTEM

As of D-Day in early 1971, detailed abstracts of all material published in The New York Times and in a wide variety of other publications will be processed into the computer. An initial data base of earlier materials is readily available from the tapes of The New York Times Index which has been in a successful computerized operation since January 1968 (this operation has served as a pilot project for the enlarged system).

Gradually earlier selected data will be incorporated into the system at a planned, orderly rate. These earlier data will be obtained from The Times Morgue clipping files which will then be retired and eventually complete phase-out of the Morgue will be achieved.

Information from earlier annual Indexes will eventually be edited for automatic system input, possibly by OCR equipment. Data which predate the actual time when the system commences operation will be drawn exclusively from The New York Times, whereas current data will come from many news sources.

The Times anticipates that future expansion of system input will include storage and retrieval of photographs and other graphic materials; bibliographic citations of relevant books and other reference materials available in The Times Reference Library; and interface with other large reference libraries and information centers using an automated system.

POTENTIAL MARKETS AND USES

The Times envisions that real-time access to its gigantic store of background information, whose depth and scope equals that of The Times' own news-gathering, will prove to be of immeasurable value to major reference and research libraries, general business services, radio and television stations and networks, public relations and advertising agencies, and individuals such as scholars, journalists, and researchers in every field of endeavor.

To assist in future planning, The Times has retained the services of Arthur D. Little, Inc. of Cambridge, Massachusetts, one of the world's most respected research organizations. It will be this company's responsibility to determine the size and scope of the potential market, evaluate the kind of response the system and the service it performs may expect from potential customers, and scientifically forecast what changes are likely to occur in the demand for the kind of service the system will be capable of producing.

[News release, Mar. 26, 1969]

NEW YORK, N.Y., March 26, 1969.—The New York Times today announced the development of The Times Information Bank, a real-time, interactive retrieval system which will make available vast resources of material to major research and reference libraries, government agencies, journalists, scholars, and other media, including radio and television networks, with speed, thoroughness and comprehensiveness.

In a news conference at the company's headquarters, Ivan Veit, a vice president of The Times, said International Business Machines Corporation and its Federal Systems Division have been retained to assist in the design and implementation of the system. Additionally, Arthur D. Little, Inc., the well-known research organization, will assist with market development.

He said, "We envision that the instantaneous accessibility of a gigantic store of background information on virtually every subject of human research and inquiry will prove to be of immeasurable value not only to major reference and research libraries, general business services and other media, but also to individuals engaged in all forms of research.

"The system will serve the news and editorial departments of The Times and shortly after becoming operational will be extended to include a wide range of customers who require rapid, comprehensive and authoritative information."

Mr. Veit noted that the first input into the retrieval system will be abstract data from The New York Times Index beginning January 1, 1968, which are already on magnetic tape. Gradually earlier selected data will be incorporated into the system at a planned, orderly rate.

He said that data which predate the actual time when the system commences full operation in early 1971 will be drawn exclusively from The New York Times, whereas current data fed into the system will come from many other sources.

"The New York Times," said the company's executive, "intends to enhance its reputation, through its information retrieval system, as one of the world's most reliable and authoritative sources of information. We feel the potential market for the services which the system will be capable of producing extends into many areas.

"For example, the services could be put to invaluable use by government agencies engaged in social research, scholars preparing such major documents as doctoral dissertations, general business services conducting research in specific areas for various clients, journalists marshalling material for books and articles. The list could be extended to include the news and public affairs departments of radio and television networks, advertising and public relations agencies, and the research arms of the many philanthropic foundations."

Future expansion of system input, he stated, will include storage and retrieval of photographs and other graphic materials, bibliographic citations of relevant books and other reference materials available in The Times Reference Library, and interface with other reference libraries and information centers using an automated system.

The system is being developed under the direction of Dr. John Rothman, director of information services at The Times.

STATEMENT OF DR. JOHN ROTHMAN AND ROBERT S. NOVEMBER, NEW YORK TIMES LIBRARY SERVICE DIVISION

Mr. NOVEMBER. Thank you very much, Mr. Chairman.

We thought we would tell you the background of our decision to undertake this ambitious new system. Dr. Rothman could describe in greater detail exactly what we are doing and how we are doing it.

We appreciate your kind words about the New York Times. We think of ourselves as a newspaper and sometimes, immodestly, as a great newspaper, where a tremendous amount of resources are devoted to gathering and presenting to our readers a variety of news.

We have almost 1,000 reporters and editors engaged in this endeavor. Back in 1851, when the New York Times was founded, the index to the Times was begun simultaneously; so we have a tradition of over a century of useful retrieval of the information in the newspaper, making it available both to ourselves and to other research users.

As everyone knows, in the last decade the technology of information has been changing very markedly and we at the New York Times felt that the time had come for us to adapt this to an information system for two major reasons.

First, and foremost, so that we could maintain our position for our readers of presenting the news comprehensively and completely, so that our reporters and our staff members themselves would have access to information in the best possible manner.

The second reason, which is also important, is that we see ourselves as not only a newspaper but as a tremendous reservoir of information, and we have been searching for ways to make that information useful, not only in the form of the newspaper you pick up in the morning or the New York Times News Service which we maintain, but in a way that would be particularly adapted to information needs.

Several years ago, therefore, a committee composed of members, the news department of the Times and the business departments, was formed under Dr. Rothman's direction to create and initiate such a system.

I think the best way to proceed is for Dr. Rothman to explain exactly what we are undertaking now.

Dr. ROTHMAN. I think probably the best way to start is by describing briefly what we had before we went into this project. The Times had four principal information facilities. The clipping library or morgue, the New York Times Index, the reference library, and a photo library.

These were completely separate, not only administratively, but they each followed their own procedures and they each had their own vocabulary. For a member of the staff to get the best of information from all four meant instituting four separate searches, going to four separate facilities on four separate floors and getting the information out in each case using the separate system.

This was a very wasteful and inefficient process. The first thing that I recommended to Times management about 4 years ago was that we try some coordination of the four facilities and then try to apply the latest available technology to making a single system out of the four.

This is in effect what this project contemplates. We are going to start with just the clipping morgue and the index but we are planning to add to this project before too long the facilities of the photo library and the reference library.

The clipping morgue is a vast repository of wealth that even we are hardly able to gage and which because of the nature of the newspaper clippings—the clipping is fragile and it deteriorates just by sitting

for a while—cannot be made accessible except to a very limited number of people.

I spend more time than I like just politely denying access to the clipping morgue to all sorts of people who claim to have some kind of privilege or need to use it.

We estimate—and this is a very rough estimate—that our clipping morgue right now contains somewhere in excess of 20 million clippings. They would roughly divide—and this is very rough, we did a hasty sampling and, of course, it changes almost from day to day—as follows: our current subject files probably contain about 3½ million clippings; our current biographical file, which consists of about 1¼ million names, contains about 4½ million clippings; and the ex-current or inactive subject and biographical files together contain somewhere around 12 million, give or take a million.

Some of the information goes back for decades. We do not know exactly how much. And some of it is very, very recent indeed. We try to process today's clipping into the files within 24 hours.

The morgue is essentially a single access system. If you are looking for material on information retrieval and a House of Representatives subcommittee you can either go to the appropriate files on information retrieval or you can go to the appropriate files on the committees of the House or you can go to personal name files, if you happen to know the names of the people involved.

In each file you may find some material that is relevant to all three along with I don't know how much material that is not relevant to any but the one. You cannot take the three separate access points, or descriptors, or clues, and put them together and get out only that material which is relevant to all three.

This is not possible in a manual system. It is possible in a computerized system. And this is possibly the largest single advantage that our information bank will give. It will make it possible for people to think of Cornell University and student demonstrations and Mr. Perkins and riots in the last week only and retrieve only that information without getting what we call noise or irrelevant material.

The New York Times Index, as Mr. November mentioned, goes back to 1851. It was an entirely separate operation. It works by having each relevant item of information in the paper abstracted and then the abstracts arranged in chronological order under subject headings which are arranged alphabetically.

The index comes out twice a month in booklets like this and then we publish a cumulative volume once a year. As I mentioned before, the index operation is entirely separate from the morgue operation, yet we handle by and large the same material or at least the bulk of the material that the morgue handles here as well.

What the future system envisages is that we take all the material that is presently processed into our morgue, which means virtually everything that is published in the Times, plus relevant background material from some four dozens or more other publications; to abstract these in great detail, to index them in great depth and to have the abstracting and indexing done by trained information specialists who will be processing this material directly into a large computer from on-line terminals.

The original clippings will be reformatted, pasted up on pages about 8½ inches by 11 inches, and microphotographed. The photo chips will be placed on microfiche, and the microfiche stored in a mechanical device which will be interfaced with the computer in such a way that any one clipping on any one fiche will be addressable directly by the computer using some equivalent of the date, page and column citation that will follow each abstract.

An inquirer will sit at a terminal, presumably a keyboard and video tube terminal, and will phrase a question in terms of descriptors, bylines, dates, the kinds of articles in which he is interested (say, editorials, letters to the editor, news analysis) and so on.

When his question has been phrased and accepted the system will search an on-line file of the abstracts, and will product on his screen in chronological order all the abstracts that are relevant to his request.

He will scan the abstracts, getting from them whatever information he desires, and then if he wishes to see any one or more of the original clippings, he will be able, by pushing an extra function key, to see an image of the clipping displayed on the screen.

That, in brief, is a description of the system as we envisage it going into operation sometime early in 1971.

There are certain exceptions that I must make. We will not be able for some time to transmit images of the clippings themselves outside of the Times headquarters building in New York because the technology of facsimile transmission is still inadequate. We expect users in remote places will either have their own microfiche file or will write or phone us for full text, or if they want something much less sophisticated, they will be able to use the New York Times microfilm.

We expect to go back almost instantly. The New York Times Index has been in a successful computer operation since January 1968. We will be able to feed the tapes of the index back to January 1968 into the system, so that we go into operation in early 1971, we will have instantly available to us more than three years of background information (although it is limited to the material processed into the index).

We will at that point, of course, have remaining for us the New York Times clipping morgue. We are planning, as rapidly as we can, consistent with control and careful selectivity, to go through all the material that will then be reposing in our morgue and we will be gradually processing it into this system.

I would not be able to even guess right now how long this process is going to take beause it means going through 20 million clippings in effect and picking out those that are most worthwhile and then gradually feeding them into the system. However, we will do this as expeditiously as we can.

We are also planning, as soon as the system is in operation, to index our photograph file in machine-readable form and feed references to that into the system. We are planning as a later step to put our reference library catalog into machine-readable form and process that into the system, so that we hope that by perhaps some time in 1975, a user will be able to inquire of the Times and obtain abstracts and copies of its material and references to other sources, bibliographic material, maps, charts, diagrams, cartoons, and other graphic material that we have.

We believe that this is going to be a staggering source of information and that most of the information in it will be available to the users in what we call real time. You query the system, you push your final question mark key, and the question is accepted and processed and within fractions of a second you start getting a responsive and relevant answer.

The technology is here. If I may make a general statement: my own feeling is that machinery is here long before we know how to use it best. We are using existing devices and in some cases the latest, refined models of existing devices.

The heart of our system is going to be an IBM 360–50–I model with probably another computer of the same kind as back-up. The descriptors and location of items will be stored on IBM 2314 disk drives. The abstracts themselves will be stored on the IBM 2321 data cell. I have only fairly recently become familiar with the capabilities of the data cell. We will be able to store in one data cell approximately 20 years worth of abstracts.

The microfiche storage and retrieval device that we are planning to use has not yet been finally selected, although we probably will have the selection made within the next week or so.

The devices that we have looked at are in operation and are perfectly capable of handling what the system wants.

For terminals we have a variety of choices. One of the terminals that we are most seriously considering is one that will be able to project on the same screen both the abstract information coming from the computer, and the video information coming from the microfiche retrieval device. That terminal, however, also has not yet been finally selected.

I think the only other thing that I might want to say now is that I am ready to try to answer any questions that you might have.

Mr. PUCINSKI. Dr. Rothman, there is no question that what you are attempting is one of the most revolutionary breakthroughs in the dissemination of information in this country. In my judgment the New York Times library, file, morgue, or whatever you want to call it, is undoubtedly the greatest warehouse of information on the humanities in this world.

You are lucky that you have a 100-year background and can draw upon 20 million clippings. You have chronicled every significant development in these 100 years, and I am very excited that the New York Times has made this decision. Surely it is a costly decision.

I am wondering if this system will ultimately be available to private subscribers, to non-New York Times people, to libraries, to Members of Congress, to whoever wants to subscribe to it, and will this be on a subscription basis or fee basis?

Dr. ROTHMAN. The answer is certainly, definitely.

I do not think that the Times could have considered undertaking a project of this magnitude requiring an investment of this kind if there were not the possibility of marketing access to this information.

The Times is a business and we do expect to have some revenue out of this system. We know that the demand for access exists as you kindly indicated in your remarks just now. We are definitely already in touch with a number of potential users and many of them have contacted us, even before we actually made the announcement.

I have certainly, in professional association with other libraries and Government agencies—Library of Congress and so on—talked about this project which has been under study for almost 4 years now.

We have had some delegations in from university libraries and government agencies including the Library of Congress. It will be possible for a subscriber to buy or rent a terminal and subscribe to access to this system on probably a range-of-fees basis. We have engaged Arthur D. Little, Inc., a very well-known and respected market research organization, to do a study of the potential market for us and one of the things that we hope to be able to determine on the basis of this study is what the range and scope of the demand is and what range and scope of services we should offer.

We have some idea of that already. We do expect that at least major central public libraries will have terminals available and that patrons of the library will be able to use it, possibly by paying a small fee to the library and the library then would subscribe to the access service on a monthly or annual basis.

Certainly we expect that most of the news media, most of the large news media, will wish to subscribe to access to this service on a fairly regular and large-scale basis.

Mr. PUCINSKI. Let me take a hypothetical situation.

A young student doing research at the University of Chicago on Sino-Soviet relations—when this system is completed I presume this young student will, by going to the library at the University of Chicago, be able to energize the questions that he has and feed them into the system.

Now, as to the feedback. Will that be a printout or how will the finished product come back to the student and what will he have in the finished product?

Dr. ROTHMAN. It depends on the kind of terminal. If it is a video terminal he will see a display of fairly detailed abstracts, summaries of news items, each one followed by its proper citation, and the abstracts will be in chronological order.

If there is a printer terminal—and I would imagine that large installations like a university library would have both—if there is a printer terminal and the student or the researcher has asked for a substantial amount of information, it will be printed out for him on a high-speed printer and he will be able to take with him computer printouts of the abstracts in their chronological order, giving proper citations.

If he wishes to have the clippings themselves, then he would have to go, as I mentioned before, to a separate store of the full text. This might be microfiche, and we are planning to offer to libraries the microfiche that we are developing for this system.

The library would have to have a reader or a reader-printer, which I understand are available at relatively low cost. He would get out the right microfiche, put it under the scanner, and obtain either an image of the clipping he can read on the screen or an electrostatic copy that he can take with him.

Where the question requires a larger volume of information he would probably—they would probably call us and we would produce either large sets of the abstracts and/or large sets of the clippings in a batch,

off line, and ship them to him at whatever rates will prevail for this service.

We expect to publish on demand all sorts of special subject accumulations. For example, if there should be an interest developing in some particular subject, importation of meat from Argentina for instance, and someone wants a retrospective grouping of material that the Times has published on this subject, we would be able to produce that presumably on a 24-hour turn-around basis.

Mr. PUCINSKI. I have several other questions but I will yield to Mr. Dellenback now and we will come back to my question.

Mr. NOVEMBER. Could I add one item to this.

As we see the technology now, it will be possible for users, primarily libraries or institutions, to use the communications stations that they have. For example, if there is a TWX machine in the library we will have a service whereby that library can subscribe and a patron can sit down at that TWX machine, ask the system for information on the House of Representatives and the output will then be abstracts printed out on that machine in the library. It is possible with the technology to have in some places video terminals which are much quicker, and in other places slower-speed printers which will serve the needs of a variety of potential subscribers.

Mr. DELLENBACK. That is part of the answer to one of the questions I was going to ask. I was going to ask about the degree of compatibility of your system. You say it is just uniquely something, that you must have your own source material on, and your own printers to make it work effectively.

You must have your own system all the way through. There is a sufficient degree of compatibility apparently—this is really a question—so that your system can feed into other types of systems and it does not require the same investment across the line, am I correct?

Dr. ROTHMAN. We expect to be able to do this to a degree. We have chosen hardware, the IBM System 360, which is the one that is most widely used for information retrieval purposes at a number of other installations.

The NASA facility uses System 360. Medlars (the National Library of Medicine) is moving over to System 360. I believe the National Agricultural Library is using it and of course the Library of Congress, Project MARC is going to be on System 360. We will provide for hardware compatibility.

Software compatibility is a different story. At this stage of the game it is very difficult to achieve and we are of course developing our own programs. (It would be very difficult to find existing computer programs that would be able to handle this system as well as any other of similar magnitude.)

However, the whole industry is striving toward greater compatibility and toward easy conversion programs. So that we expect in the future to provide the convertibility and of course we have that very much in mind.

We are checking—in the course of developing the system, we have been checking and are continuing to check—on what is being done at other large installations of which I have only mentioned three or four.

I have spent a great deal of time over the last years doing this. I am continually in touch with other large information users and large

information systems. We hope that everyone will be sufficiently aware or will be made aware of the necessity to provide software compatibility.

Mr. DELLENBACK. Without seeking any special business information which you consider confidential and is not what I am seeking to reach for, can you give us some idea about costs involved, either cost of installation or cost of utilization anticipated?

Mr. NOVEMBER. We have already said in response to this question, that is a several million-dollar investment. I think if that is agreeable with you we would like to leave it at that.

Mr. DELLENBACK. So far as the utilization, you talk about working on a subscription basis and then perhaps a major library could have an individual user fee within that basic subscription. Can you give us any idea of what you anticipate these costs might be?

Mr. NOVEMBER. This is a preliminary estimate on our part. As Dr. Rothman mentioned we have a consultant, Arthur D. Little Co., to help us formulate this to the need of our users. Our current estimate of the subscription charge would vary from $100 a month to $2,000 a month, depending on the kinds of service the user wants. Obviously that figure will have to be clarified before 1971. Getting back to the previous question I would like to emphasize again that we think of this as a tremendous information base that we hope that frankly almost every library will use.

For that reason, we are working to make it as easy as possible for them to do so. In other words, the output terminals should be those which would be available not only with our system but with other systems. So, one of our objectives is to have this hardware compatibility on the output end.

I think Dr. Rothman did not emphasize enough, also, that the system is being designed so it will be easily usable by normal information seekers, whether reporters or graduate students or undergraduates or Congressional assistants.

It is a regular English language system. Instead of speaking to a librarian and saying, "I want information," you sit at the typewriter and use the English language to get the information out.

Every attempt is being made to make this as widely usable as possible.

Dr. ROTHMAN. I would like to add one thing to this.

We are designing this system for two primary purposes. One is to serve our own staff better and the other is to make the information facilities as accessible as possible, to anyone who has need for this service.

Our own staff is likely to be extremely demanding. A reporter who is working against a deadline is not going to be able to go through very elaborate coding and he certainly is far from being an electronics engineer, so we are designing this system in such a way that he will be able to use it as easily as—and perhaps more easily than—he is presently using the morgue system, such as it is.

My own feeling is that if it is right for this person working under those circumstances, then this system will be relatively easy to use for almost anyone who might have need for this kind of information.

Mr. DELLENBACK. May I ask a further followup question on the matter of cost? Recognizing that within the range, the $100 to $2,000

a month range, based on types of use and how much time and so on, are you able to give us any unit down to the student who wants to use it for 5 minutes or 1 hour, how much is it apt to cost him to make use of the full resources?

Mr. NOVEMBER. I think we have not addressed ourselves to that question primarily because we expect most students to do this through their libraries, whether university library or public libraries.

Mr. DELLENBACK. I assume your cost of saving $100 to $2,000 a month is in fact based on some units of use. For $100 one would get x minutes or x hours of utilization or for $2,000 you would get 20 x or so on.

Mr. NOVEMBER. We don't see that as individual use of the machine. We see that as the libraries' use.

Mr. DELLENBACK. I am just trying to get some idea, if we talk way beyond your system and talk about a nationwide retrieval system—I do not have any concept at the moment of what realm of cost we are talking in. If you stay with your concept of $100 as a monthly fee or $2,000. or any interim fee you want to use, what would come for that figure? How much would one be entitled to get in the way of measurement of use for $100 a month?

Mr. NOVEMBER. Well, I was going to say, let us start at the other end.

Mr. DELLENBACK. $2,000.

Dr. ROTHMAN. If you will be good enough not to hold us to the precise figures.

Mr. DELLENBACK. We will consider this not at all binding legally and just to give us some road in. What do your calculations at this time indicate?

Dr. ROTHMAN. We are assuming for example a major publisher or television network that is going to want access to this instantaneously on say a 20-hour-a-day, seven-day-a-week basis would pay the maximum figure, on a monthly basis, and for that would be entitled to, I would imagine, somewhere between 10 and 20 questions and answers a day.

This is pure guesswork.

If they exceed that then there might be some surcharge, depending on the excess.

It may also turn out that for the large-volume users we charge a flat fee for as many turns as they might want at the device.

Mr. DELLENBACK. Pushing the stick then that you have given, let us assume that this user who might ask 10 questions a day would pay $2,000 a month for the service. This might be on, what, a 5-day week.

Dr. ROTHMAN. No. 7.

Mr. DELLENBACK. So, we have 300 questions which, in effect, might yield a price of $2,000?

Dr. ROTHMAN. Something like that.

Mr. DELLENBACK. We are oversimplifying this but I am trying to get a grasp on some unit of measurement.

Dr. ROTHMAN. When you are talking about the other end, the student or individual library patron, I would like to perhaps just draw an analogy. Many libraries now have electrostatic printers, big Xerox machines, available.

They either buy these machines or rent them from Xerox paying the standard going rate. Then, depending on volume, they charge individual patrons a quarter or 50 cents for use of the machine.

It seems to me that what we will probably do for public and university libraries is much the same sort of thing. The library will pay, depending on its anticipated use, and in turn it will charge its patrons. And the library being a public institution will probably charge just enough to defray its costs and we may have to have some agreement whereby on an annual basis we revise the charge depending on volume.

This is much the way that the scientific services, as far as I understand, now charge the industrial users. The one that I am most familiar with is Chemical Abstract Service. They charge for their computerized or publications service on the basis of the number of users at a given installation.

So, if you have a relatively small company, with just three scientists using it, they pay considerably less than would Du Pont in Wilmington. I imagine that we will follow much that kind of pattern.

Mr. DELLENBACK. Again with all the caveats that you want and I am willing to give you all the ways out that you want, if we can stay with your figures, if it means 10 to 20 questions a day and $2,000, this would mean somewhere between $3 and $7 a question.

Somewhere in this area is what I am thinking about, if my mathematics work out correctly, off the top of my head, which tells me something about the average research student in a library; he might find that he does not use it the way a scientist or the way somebody really researching a project in depth would be able to say, "This question is important, that is a minimal cost," but a student who is writing a thesis on ancient Greek pottery might think hard and fast before he started to feed 10 or 20 questions into a machine at that unit of cost.

Dr. ROTHMAN. Let me make a distinction. To some extent it depends on what kind of response they want. If they are sitting at a cathode ray tube terminal and are getting the answers displayed back, this is more expensive than if they write to us and say, "Print it out for us at 2:30 in the morning when your device is not used any other way and send it to me parcel post. I have plenty of time."

So, I don't know at this point whether we have to base our charges on the number of items requested, the mode in which they are supplied, the number of lines printed out or displayed.

Chances are that all these elements will enter into it and this is precisely the reason why we have retained Arthur D. Little as a consultant, to help us frame the proper basis for these charges.

Mr. NOVEMBER. I would like to add that to some extent the variety of the output is a variety of degrees to which you are making the computer do the work that you would otherwise have to spend a great deal of time and effort to do.

We do anticipate, for example, that we would not have to charge at all if you were to call up as a graduate student and say, "Tell me what stories you have run on archeology in Greece in the last year."

Our computer could give him those citations and if the person had the time he could then go and look up each of those stories. We would envision many small libraries would have an intermediate step of TWX machine to call the New York Times, type the question, and at relatively slow speed get back some of the answers.

Our charge for that would be less than having access to instantaneous use of information on the cathode ray tube.

Mr. DELLENBACK. There is a series of questions that would be interesting to follow but one more that is broad in scope—is there anything you can tell us as to the major problems you face—oversimplify your answers as much as you like—in trying to set up a system of knowledge retrieval when you want to gather together basic raw materials, somehow digest it in digestable form and then make it available to somebody else.

What are the major areas of problems that you face? Costs? Getting the right hardware, developing the software, what?

Dr. ROTHMAN. One of the major problems is behind us and that was getting management approval for this project.

Mr. DELLENBACK. There are no senior officers present with you?

Dr. ROTHMAN. I am joking.

I touched on the answer to this before. I think that hardware is not a major problem. The hardware is here. Also the ability to program is here. The programing language is available to us. I think it is sufficiently capable that we can do almost anything that we have to.

I would say the biggest problem is to define precisely what we want out of the system and how to structure our basic vocabulary to get at it. And then to translate that into systems specifications that the systems analysts can work with. I would think that that is the biggest problem.

I have done considerable reading and work in this field, also outside of my work for the Times. I have been chairman of a subcommittee that has drafted a standard of indexing which was recently published as the "U.S. Standards Institute Basic Criteria for Indexes."

There is the problem of semantics, of taking a piece of information, a document, text, and describing it in such a way that whoever wants it at whatever time and in whatever context can get that document out of the file, and no other. That is the biggest problem.

I think that—incidentally, I am optimistic enough to say this—I think that for our system we have licked it.

Mr. NOVEMBER. I was going to agree with Dr. Rothman. I think the major problem has been an intellectual system design problem. The opportunity of somehow using the hardware and getting all this information back in the way you want it is very dazzling.

The real problem has been to create an approach—a framework—which will make that practical.

I think it is only fair to correct any erroneous impression. I think the management of the Times has been overanxious in the other direction.

They have been saying to us, "What are you waiting for? Why don't you have a system that we can use internally and externally?"

The problem is: Does the information group have a system that we are confident will serve the users? To create that framework is very difficult. Cost is a problem. And it is only because we are confident that we have a system that will be efficient, both as a resource for the rest of the world and for our own use that we are therefore able to justify the expenditure which this entails.

Mr. PUCINSKI. Your testimony is invaluable to this committee for

many reasons. One, because I frankly believe the biggest problem is methodology. I don't think we are doing enough in this country.

Our various agencies are spending all kinds of money on research but they are really spending very little money on addressing themselves to the question of what we want out of an information-retrieval system.

What language do we use? How do we cross index? What methodology is best? The Weizman Institute of Science in Israel is doing some exciting research in this field. There are a few others that are researching the problem but I must say I am very disappointed that those who control the purse strings in Government have really not had the imagination to fund reserves in this field to any great extent to help people and organizations such as the New York Times.

The Science Information Service director was before the committee yesterday and referred to all the things they are doing. But I believe they are really not doing very much in this direction.

The other thing I was impressed with is the initiative of private enterprise. You have taken on this very costly project and you are not worried that maybe tomorrow it is going to be obsolete. You are taking a chance today.

This is certainly in sharp contrast to the testimony yesterday, when the Director of the Science Information Service indicated his office could not support a National Information System because new technology may make it obsolete. This is the same attitude that was expressed 100 years ago by the Director of the U.S. Patent Office when he suggested we shut down the Patent Office because there was nothing more to invent.

As I say, thank God for private enterprise. Obviously, if we were to leave this project to existing Government agencies to do, it would never get done. I take it your system is an evolutionary system, one that will adapt itself to changing technology, and the various advances that obviously will be made in hardware as we move along. Is that a safe assumption?

Mr. NOVEMBER. Yes, sir.

Mr. PUCINSKI. And it is safe to assume after the huge amount of money that you are going to put into this sytsem, you are not going to swap it for another system in 5 years. Rather, you are going to build on the system you have.

One question comes up. Had you waited, had you not committed yourself, as you did, to a program which you hope to have in operation by 1971 and fully operational with the full 20 million clippings by 1975, had you waited as so many people have advised this committee to wait, do you have any idea what would be the increase in cost with every month that you delayed development of this operation or every year that you delayed it?

Dr. ROTHMAN. No, I would not be able to say, because there are several kinds of costs. One is the price, hardware and software, in developing the system. There is also the cost which we feared more actually—and by "we" I not only mean us in the team that was working on the project, but I am quite sure our management as well—is the cost of not doing something that would be worthwhile, and that would supply a need and bring in an apropriate amount of revenue.

And the concomitant fear that one of our competitors will move into the vacuum instead.

Mr. PUCINSKI. Of course, you have one built-in guarantee, one indestructible safeguard, and that is that you have 100 years of knowledge and information stored in your New York Times library. As fas as I know; no one else is capable of catching up with you on the score.

Ultimately the Library of Congress may be computerized, as well as other sources of information. Having had a little experience in this field, it would be my judgment that it would be almost impossible for anyone to catch up with your built-in advantage of having your own warehouse of knowledge in the humanities. As far as I know, no one else can equal it.

Dr. ROTHMAN. That is true, but we have found that most of the searches for information will probably go back somewhere between 5 and 10 years and that the searches for information going further back may be relatively rare. This is a relative thing. So that possibly some other information vendor who does not have quite the file that the Times has, but who has a respectable file, nevertheless, who is out before us, would be able to preempt a good portion of the market.

We face this in microcosm with the index. The index goes back to 1851 and it covers the New York Times. Yet there are any number of published current information services around that give us a run for our money.

Mr. NOVEMBER. We did not aproach the question as to what would be the cost of waiting. We were a little bit the other way around. If now we have a system and technology that can do this for us—and again I would like to emphasize the benefit we feel for our own staff of being able to get information quickly and comprehensively—we should not wait.

The second item would be what is economically feasible at this time to do.

Mr. PUCINSKI. Dr. Rothman, you said the system at the New York Times was developed by a committee which you headed. How long had this committee functioned?

Dr. ROTHMAN. The committee was formed in June 1966. In back of it is about a year and one-half of active work that I put in pretty much by myself. The committee still functions. The committee consists of Mr. November and myself, the former chief librarian of the Times, who is now general services manager, the Times systems manager, and two relatively recent employees of the Times who have joined the Times as part of the information service staff.

The other member of the committee is a representative of the news department, the assistant metropolitan editor, who kind of represents our customer interest.

Mr. PUCINSKI. The reason I ask that question is that the legislation before us, H.R. 8809, calls for the establishment of a national information retrieval system. It describes in general what it should be like.

Perhaps it would be a good idea to rewrite this legislation and establish a Presidential commission that would be charged with the development of this system. I am not sure, but perhaps that might be a more realistic way of proceeding on a national scale with the work you have already undertaken.

In effect, you really have a prototype of what we are talking about in science information. You have illustrated what can be done in this field. I am grateful to you for bringing us up to date on the fact that the technology is here, that the hardware is here, that the methodology is here. While the New York Times will concern itself essentially with largely disseminating information on the humanities, the bill we are discussing calls for setting up a system to retrieve information in scientific and technical research.

I am very grateful to you for this testimony.

Mr. Ruth, do you have any questions?

Mr. RUTH. No questions.

Mr. PUCINSKI. I want to make one observation, that the New York Times of 1969 is a far cry from that of 1920. I once came across an editorial in my research on the New York Times in 1920 in which it was suggested that Dr. Goddard be fired. In effect, the editorial called him an imbecile. It stated that anyone who would suggest a rocket can be launched out of the force of gravity and then propel itself into outer space and around the moon must be completely out of his mind, and any further expenditures on that kind of project is just a waste of taxpayers' money.

That was a very fine editorial in the New York Times in 1920. I am very happy to know in 1969 there is considerably different thinking at the New York Times.

Dr. ROTHMAN. May I respond to that? I have been asked by someone—and I don't know whether he was trying to be funny or whether he was being incredibly naive—whether in this system we are going to go back and correct incorrect material.

The answer is no.

Mr. DELLENBACK. I am not sure that was a gracious way to close, Mr. Chairman. May I ask a couple more questions?

You indicate it has taken you some 4 years, in effect, of your time so far, Dr. Rothman, and the committee something less than that, but you have been on this for 4 years working toward this particular issue.

If I understand your testimony correctly, you indicated that the prime body of knowledge you are going to be working with is Times knowledge, for a period of time after which you will reach out beyond, but you begin with your own, and this is a relative word, relatively limited source of raw material, a very substantial source but relatively limited.

Do you have any concept of what it might take you in time to get ready even on the basis of your research that you made to date on this system, if we were to charge you with the responsibility of developing an information retrieval system that would embrace all branches of scientific research, covering every possible field of scientific research, not just within the United States of America, but on an international basis?

Would you be ready to start on that immediately or would it take you a little time to get ready for it?

Dr. ROTHMAN. I think it would take more than a little bit of time. Let me respond to that in several ways. How long it takes to get a system ready depends to some extent on what is there when you start.

In our case there were, as I say, four separate facilities that had been in operation for some time and that were, all of them, entirely manual. There was no machinery used of any kind.

I have been working for the Times for almost 23 years. So I started out with a very substantial basis of knowledge of what the Times had in those systems and how it was being processed. I had a fairly good idea of how it was being used.

Now, to draw the parallel to this system that you are questioning me about. Someone would have to make a study of what is now being done, how it is being done, how it is being used, and how should it ultimately be used.

How long that takes depends to some extent on how knowledgeable this person is of the field or fields and how much of the data already is available for him to see.

Mr. DELLENBACK. Since there is an immense wealth of scientific papers that is of record already across the scope of the Nation and the world, we can imagine that there would be a tremendous problem, if you would reach backward at all to try to put this in form.

Would you tend, if you were approaching that sort of problem, realizing its complexity and breadth, to try to create in the original instance a program to cover the entirety of it or would you tend to approach it from the point of putting together pieces of it [illegible] and then trying to weld the pieces together?

Dr. ROTHMAN. I am going to answer that out of the depth of my ignorance of the natural and physical sciences, which is [illegible] but because many of the physical scientists that already have [illegible] systems going tried to tell us how to go ahead and process [illegible]

I would imagine that the best way to go would be [illegible] and gradually. This is in some sense already being done. [illegible] limited knowledge there is already some interchange taking place between the biological sciences and the medical sciences [illegible] the National Library of Medicine and Biological Abstracts [illegible] is very good.

There is some interchange between the Chemical Abstracts Service and the National Library of Medicine, [illegible] and the National Agricultural Library [illegible] It would seem to me that [illegible] perhaps [illegible] to start. [illegible]

[illegible] for example, Psychological Abstracts [illegible] with the [illegible] the National Aeronautics and Space Administration with the Engineers Joint Council and the Department of Defense with the [illegible] Council.

There is a good deal of cooperation [illegible] that the way to start [illegible]

This is what we were talking about before [illegible] extent this means compatibility of people and [illegible] and systems, just as well as compatibility of hardware and software. I would urge you to use the prestige and influence of the Congress to encourage these scholarly and professional organizations to cooperate

and perhaps that way build a National System or at least that way begin to build National Information System.

Mr. DELLENBACK. While I oversimplify, you would stress cooperation, you would stress moving ahead firmly, but you would make haste slowly.

Dr. ROTHMAN. That does not necessarily follow. I think that the cooperation—you realize, incidentally, now I am expressing personal opinion—the cooperation between two related fields and two related institutions or services, that need not be made slowly. I think that should be done as quickly as possible.

Mr. DELLENBACK. I said, press ahead firmly, but when you talk about the total system, you would make your haste somewhat slowly in trying to get a total system. Is that correct?

Dr. ROTHMAN. Yes, sir; I would advise that. The complexities are such and the fields are so vast and the problems are tremendous, I think the best way to solve them is to approach them piecemeal rather than totally.

Mr. DELLENBACK. Thank you very much.

Mr. PUCINSKI. Just so there is no misunderstanding, in the event my colleague is trying to lay the foundation for delaying action on the legislation for us——

Mr. DELLENBACK. Not a bit, Mr. Chairman.

Mr. PUCINSKI. He asked the right questions. And I just want to make sure that we are all talking about the same thing. The New York Times has made a very important decision. It made a decision that at this point in time it needs a better way of disseminating the vast amount of information that it now has and has accumulated over the last 100 years. Any further delay in the development of this system would merely further complicate its development.

After sputnik it became very apparent to the Members of Congress and to the scholars and researchers of this country, that there was a very serious gap in the dissemination of scientific data.

The Office of Sciences Information Services was established to close this gap.

For 11 years it has been going through a lot of money, but I have not been able to find any evidence that it has moved forward in successfully managing the dissemination of information. Eight years ago we introduced this legislation, which would commit this Nation to the establishment of a national information retrieval system in the sciences.

This legislation clearly recognizes that it is best to move within the scientific disciplines. The American Chemical Society is setting up its own retrieval system in chemistry. The various engineering societies are setting up their own systems; various other segments of the sciences are setting up their own systems.

Because we know that there has to be a certain amount of cross-breeding of information among the disciplines, we have proposed that all of these ultimately—not tomorrow, not day after tomorrow—but ultimately be tied together on a coaxial cable, a network which will make this information available to scholars and researchers. My judgment is that in due time the New York Times system may very well become one of the strong components of a national system, on a contractual basis.

H.R. 8809 provides that the National Information System can enter into a contract with a system like the New York Times to feed its information into the system.

Would you care to express an opinion as to whether or not this country can wait any longer for a commitment to try to bring some sort of order to this dissemination of scientific information?

H.R. 8809 specifically forbids the duplication by the Government of existing systems, because we are not trying to put anybody out of business or make the Government compete with private or other organizations. All we are trying to do is put all of these systems into some sort of an orderly national system where their information can be readily available to whoever needs it.

Do you care to express an opinion?

Dr. ROTHMAN. A limited opinion, limited by the fact I am not in the natural sciences and have never been a user of any of the systems. My only familiarity with them has been to see what individual institutions have done in the automatic information retrieval field, to see what I could beg, borrow or steal from them for our system.

You used a term in your discussion just now that rang a very familiar bell with me and that was the term network. My own idea is that a single monolithic service coming out of a single computerized system covering all possible disciplines, is probably beyond the technology of the immediate future.

There is also this to be said about this single, giant, comprehensive facility. This is very good when someone wants to browse the field or fields. It is not that good when someone has a very specific question aimed at a special item.

These are the two main conflicting user interests, the specialist who has the one thing in mind that he wants as against the person who does not know exactly what he wants or who wants a retrospective survey.

So I would favor the network concept, whereby someone searches one system, and if he does not find there what he wants, that system steers him elsewhere—What you really want is related information—and he is switched automatically, if possible, into a compatible system having cognate information.

I think that the main search tools, dictionaries, thesauruses, instructions, the programs, should be compatible, definitely. Whether it would be absolutely necessary to house all the information in a single warehouse or to make it accessible——

Mr. PUCINSKI. The Soviets have a single warehouse and our studies have indicated that this is not the best way to do it. We, on the other hand, encourage a network of all types of selected information.

We don't want to disturb these systems or put them out of business or compete with them. All we want to do is make sure what they are doing is available readily to whoever needs that information.

Dr. ROTHMAN. On a network basis, I would definitely be in favor of it. We have talked at times, among the members of the committee and with our management, about the possibility of eventually linking our system to related systems. These are not in the planning stage. I don't know whether they will within my lifetime come to fruition.

This depends on many factors that are now beyond our control, t[illegible] are not now within our control. But I would think for instance, th[illegible]

should be possible for someone to browse the New York Public Library's catalog or the main catalog of the Library of Congress or perhaps a union catalog, for information and then find that some of this is in the archives of the New York Times.

And then flip a switch and switch over into our system and search it. And conversely it should be possible for someone to search the Times for information, find there is background information, possibly at the New York Public Library, and automatically switch over to their catalog.

Mr. PUCINSKI. Don't you think, Doctor, that there is an urgency to legislation such as is before this committee, although perhaps not in its present form. I am not wedded to the language and the provisions of this particular bill. But what I do believe very strongly and I would like to get your reaction to this—at this point, there ought to be at least a clearinghouse of information to cope with the proliferation of systems that are being developed all over the country. Millions upon millions of dollars, billions of dollars are being invested in information retrieval. Unless given some guidelines in compatibility, we will never be able to tie these systems together. And it seems to me that the key word in 1969 ought to be "compatibility."

We do not want to tell people the manner in which they are supposed to run their system. You made an excellent presentation today, and I am impressed with the way you are moving in this system, and I don't want somebody looking over your shoulder and telling you how to do it.

We leave that to private enterprise and to your good judgment. But we would like you to know what others are doing in terms of compatibility, so that as you spend the millions of dollars that you are spending, by 1975 you will not find that you are all by yourself unable to tie into any other system. We did that in the poverty program. Millions of dollars were spent on setting up some 127 information centers around the country, and then it was discovered that they do not fit.

Today they are totally useless to us. Because they cannot work, they cannot integrate with one another in terms of information exchange. For that reason, I do feel there is a certain urgency in the subject we are talking about.

If for no other reason than with the huge breakthroughs that are being made all over the country, some which should provide a clearinghouse of information so we can make some effort toward compatibility. Is there any basis for that statement, Doctor?

Dr. ROTHMAN. Again speaking as a private individual, I would endorse that wholeheartedly; yes.

Mr. DELLENBACK. Would you yield, Mr. Chairman?

I think this would be helpful in clarification of the chairman's objectives. You are not talking about a system whereby the Government would do the retrieval?

Mr. PUCINSKI. Oh, no.

Mr. DELLENBACK. You are not talking about a system where the Government would do the cataloging and the digesting and the translating but would merely be a service organization that would help other groups like the New York Times, private enterprise in some

situations, or chemical society groups or other disciplines, in their efforts to set up their own systems. We would merely stand back, exchange information, act as a clearinghouse and try to aid voluntary efforts rather than to do it ourselves.

Mr. PUCINSKI. Absolutely. This bill contains very simple language. For example:

> In order to avoid unnecessary and costly duplication in scientific research and to assure quick assess to, and a constant inventory of, all science research data, such a nationwide system shall include close voluntary cooperation with, and utilization of, on a contract basis wherever practicable, all existing science research data processing and information retrieval facilities in the United States and its possessions including Government agencies, private and public universities, private and public laboratories and libraries, abstracting societies, professional organizations dealing with specific scientific disciplines and any other facilities dealing with dissemination of scientific research information.
>
> The purpose of this Act is to implement, not substitute, existing information retrieval facilities. Therefore, it is specifically prohibited under this Act for the National System to establish any Government-owned or operated science research data processing or information retrieval facility where such a facility already exists under either private or public ownership. The National System shall use every voluntary means to arrange for an orderly cataloging, digesting, and translating, with the aid of electronic devices, if necessary, of all scientific research data produced in the United States or available in the United States from other nations and through the national science research data processing and information retrieval system, make such information readily available to any scientist or researcher, either privately, publicly, or self-employed, through an appropriate communications network. The national system shall arrange for appropriate financial payment for all science data provided into the national system by private source.

All we are saying here is that neither the New York Times nor the Sperry Rand Corp. nor anyone else in this country is in a position at the present to try to provide the tie strings, the tie rods, to put all these systems on a common network.

Our purpose is to bring together the private resources and help them set up communications among each other. That is the purpose of this bill.

Mr. DELLENBACK. The purpose of the bill, then, for clarification again, Mr. Chairman, is not actually to create any system to do this actually. It is merely to be an aid in and supplement to voluntary private systems which are attempting to come into being and creation.

Mr. PUCINSKI. I think that is properly and correctly stated. The gentleman is correct.

Mr. RUTH. As I listen to this gentleman talk and listen to you, there is no question about the value of the information retrieval in Government and everywhere else, where you learn that time is money. But the thing I was most impressed with is when we talk about a commission or committee which needs to get to the bottom of this.

Before we get to the language of the bill, I am thinking in terms of the study of the possibility to integrate the current methods, to define the purpose of the system, and investigate both limitations and possibility of the system and to evaluate it both as to cost and to timing.

I don't see how you can get into something like this without a commission or committee or somebody doing this research prior to the time that we try to put the language of the bill into effect.

Mr. PUCINSKI. Mr. Ruth, I have been here long enough to know if you want to kill something, appoint a committee to study it.

Mr. DELLENBACK. So they created a committee which worked on it for 4 years and then moved.

Mr. PUCINSKI. We discussed that earlier today.

I am perfectly willing to consider rewriting this legislation. Perhaps Dr. Carter gave us a good suggestion yesterday when he suggested that rather than try to define the system as we do in the H.R. 8809, perhaps we ought to establish a national commission that will then spell out the system.

I am not too sure that this is not a better way to do it.

Mr. RUTH. As you point out, you have been here a long time and I just got here but since we have the committee system of Government in the House of Representatives, I am a little suprised to hear you say we should not put a project like this in a committee. I am completely aware it has been said if you want to lose something, give it to the committee, but let us give it to the proper committee and let us not be ineffective because we are afraid we are going to lose it.

Mr. PUCINSKI. I don't want to disscuss the procedures in Congress, but the longer you stay here, the more you find out how frustrating the committee system can be. Mr. Bell?

Mr. BELL. No comment, except a camel is a horse put together by a committee.

Mr. PUCINSKI. Dr. Rothman, we are really grateful to you gentlemen for your testimony. You have given us an insight into a system that is now under development. To that extent I think we are all better qualified to study the legislation before us and I am very grateful to you. I congratulate the New York Times for again, as it has for 100 years, taking the initiative. This is a major breakthrough for America. I do not think the American people are aware yet what you are doing. But they will be aware when you get started. Thank you very much.